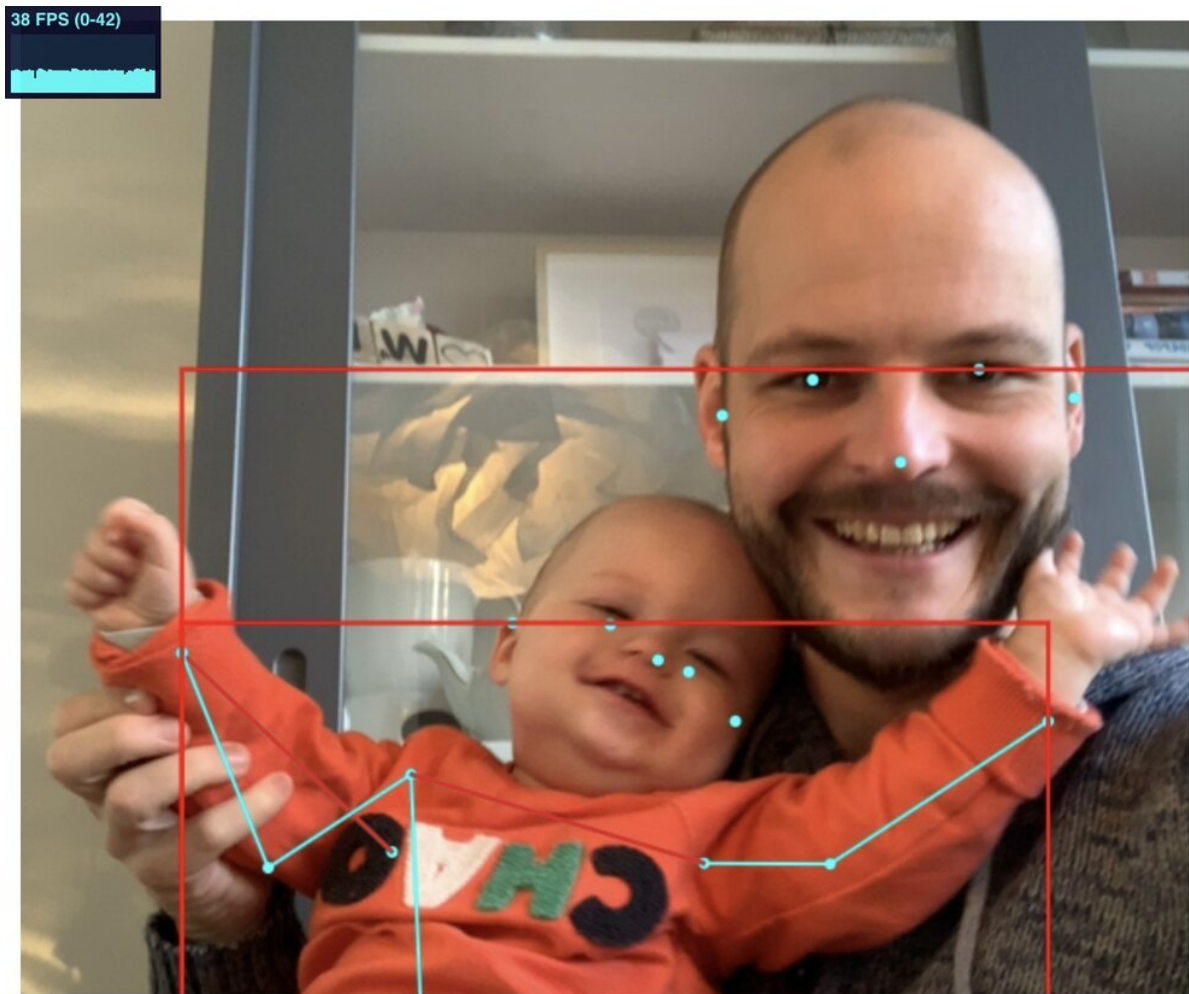


Realtime herkenning van menselijke houdingen via computervisie

TensorFlow en PoseNet gebruiken op een videofeed

Willem L. Middelkoop

1 dec. 2019



Voor een spannend nieuw project experimenteer ik met computervisie met behulp van TensorFlow. Ik wilde realtime detectie van menselijke houdingen bereiken om interactieve videoprojecties en games aan te sturen. Tijd om in de wereld van machine learning, tensors en computervisie te duiken!

Computer vision

Echt spannend is het interdisciplinaire wetenschappelijke veld van computer vision, dat zich bezighoudt met hoe computers een hoog niveau van begrip kunnen krijgen van

digitale beelden en videofeeds. Het omvat uitdagingen om digitale beelden te verkrijgen, te verwerken en te analyseren. Uiteindelijk wil je dat de computer beslissingen neemt op basis van zijn begrip van de (complexe) wereld om hem heen.

TensorFlow

TensorFlow is een end-to-end open source platform voor machine learning, oorspronkelijk ontwikkeld door Google. Het is gebaseerd op wetenschappelijk werk van verschillende wetenschappers en is sindsdien uitgegroeid tot een stabiel en robuust ecosysteem van tools, bibliotheken en community-resources.

De naam "TensorFlow" is afgeleid van het concept van een "stroom van tensoren". Een tensor kan het best worden omschreven als een "ding", zoals iets dat de computer herkende - of probeert te analyseren. Technisch gezien is een tensor een multidimensionale array met numerieke waarden. Dit stelt de computer in staat om verschillende "dingen" te vergelijken, te zien hoe hun kenmerken op elkaar lijken (of verschillen).

De grootste uitdaging bij realtime beeldverwerking is om de complexiteit van het beeld te reduceren tot verwerkbare brokken die kunnen worden uitgedrukt als tensoren. Meestal omvat dit vele stappen die het computer vision-algoritme vormen. Stappen zoals het herkennen van een vorm, het bijsnijden in die specifieke vorm, het formaat ervan wijzigen, de kleur ervan verwijderen, de contouren vergelijken, enz. Tensoren stromen door deze stappen, van 'ongeanalyseerd' naar 'herkend', vandaar de naam TensorFlow.

Verwerkingskracht

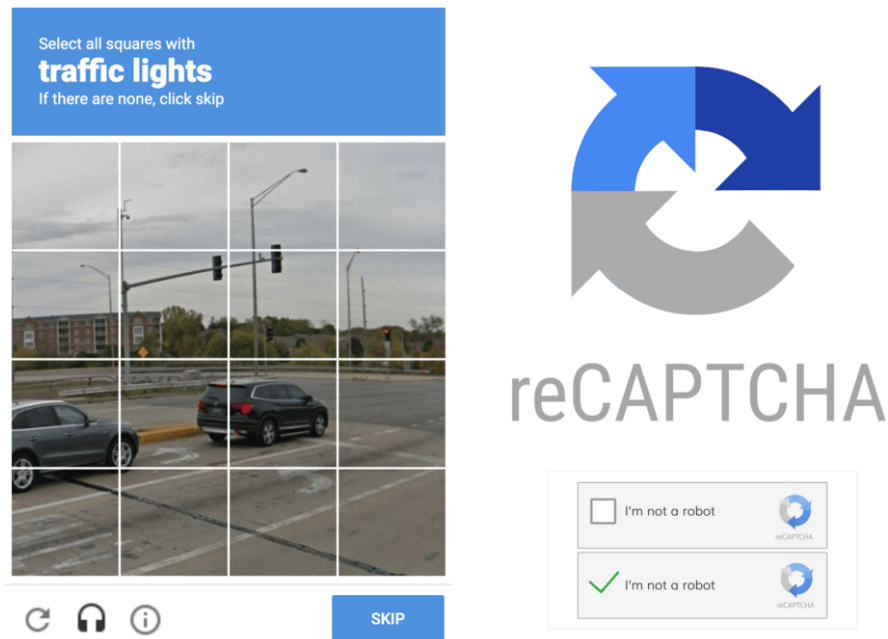
Het mooie van moderne machine learning met TensorFlow is dat je over het algemeen geen (zeer) dure hardware nodig hebt om resultaten te bereiken. Dit komt omdat het "zware werk" wordt gedaan bij het trainen van machine learning-modellen, die je kunt voorstellen als "referentietabellen". Deze modellen worden gemaakt door het analyseren van bekende datasets, waarbij de computer het 'antwoord' krijgt en de kenmerken van het onderwerp bepaalt. Het trainen van modellen vereist tijd, verwerkingskracht en trainingsdata.

Trainingsdata

Google deed iets heel slims, het verdeelde de training van zijn machine learning-modellen onder ons allemaal. Ja, inclusief jou!

Recaptcha

De kans is groot dat je "reCAPTCHA" of "Recaptcha" bent tegengekomen. Een visuele uitdaging die je moet voltooien wanneer je iets online bestelt of je ergens probeert te registreren. Deze uitdaging biedt een bescherming tegen (spam)bots voor webshop- en website-eigenaren. Het doel is om mensen te onderscheiden van bots. Dit doet het door (complexe) afbeeldingen aan mensen te tonen en hen te vragen objecten in deze afbeeldingen te identificeren. Zoals verkeersborden, kruispunten, auto's, enz. [Google gebruikt deze menselijke feedback](#) om zijn modellen te trainen.

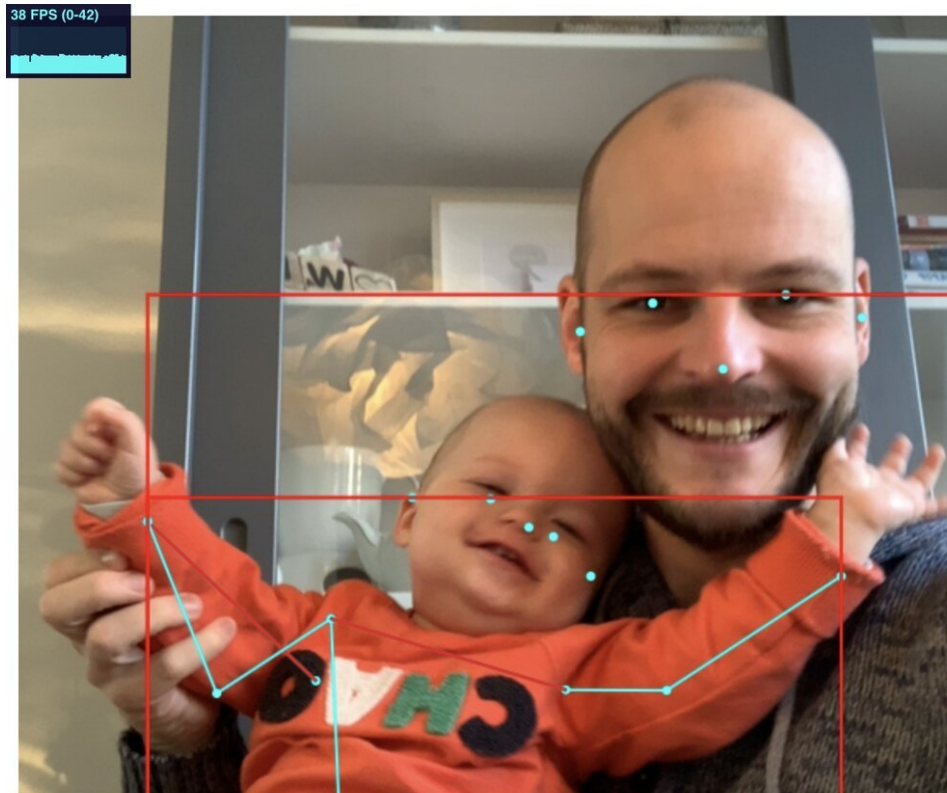


ReCaptcha: je traint al jaren Google's AI!

Het mooie is dat veel van Google's bevindingen openbaar beschikbaar zijn, er zijn veel interessante open source projecten die ons - simpele ontwikkelaars - in staat stellen om geweldige dingen te doen met machine learning.

Menselijke poseherkenning met: PoseNet

Een van deze beschikbare machine learning-modellen is [PoseNet](#). Het is een visueel model dat kan worden gebruikt om de pose van een persoon in een afbeelding of video te schatten. Dit wordt gedaan door te bepalen waar de belangrijkste lichaamsgewrichten zich bevinden, zoals je ellebogen, handen, heupen, knieën, enkels, enz.

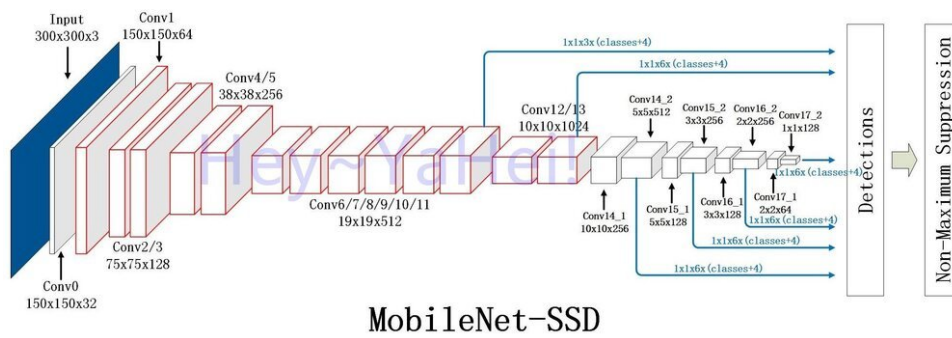


Ik en mini-ik worden herkend door het PoseNet-model

Single Shot MultiBox Detector (SSD) algoritme

De Single Shot MultiBox Detector is een populair algoritme om objecten in afbeeldingen en videofeeds te detecteren. Het is populair vanwege zijn snelheid. Het principe is relatief eenvoudig te begrijpen: het algoritme probeert de beeldcomplexiteit zo snel mogelijk te verminderen, door alleen diepe analyses uit te voeren op verkleinde, vereenvoudigde feature maps.

Deze aanpak werkt het beste als je grote objecten wilt detecteren, omdat ze snel opvallen: het genereren van voldoende high-level features om al vroeg in het algoritme voorspellingen te doen. Dit is geweldig voor lichaamsdetectie, wat een van de eerste stappen is wanneer je menselijke poses wilt bepalen (eerst kijken *waar* een persoon is, dan *wat* zijn/haar pose is).



MobileNet-SSD algoritme (image: hey-yahei.cn)

Beeld- en persoonsherkenning is iets waar de mensen uit [China](#) behoorlijk goed in zijn, om verschillende redenen. Hun algoritmen presteren goed op relatief zwakke hardware. Dit maakt computer vision op allerlei apparaten mogelijk om goed te werken, in realtime!

Coolle use cases

Het kunnen detecteren van poses in realtime maakt allerlei coole dingen mogelijk. Je kunt - letterlijk - de gamecontroller of input worden voor een interactieve kunstinstallatie!

PomPom spiegel

Met 928 bolvormige faux fur puffs is deze [kunst installatie van Daniel Rozin](#) geweldig. Het beeldhouwwerk wordt bestuurd door honderden motoren die silhouetten van kijkers bouwen met behulp van computer vision.



PomPom faux puff spiegel van Daniel Rozin

Je in een vogel veranderen: The Treachery of Sanctuary

Mensen in staat stellen zichzelf even te verliezen was alleen mogelijk door verschillende disciplines samen te brengen. Deze geweldige [interactieve kunst installatie geregisseerd door Chris Milk](#) is een gigantisch drieluik dat kijkers door verschillende stadia van de vlucht voert met behulp van computer vision.



Mensen veranderen in vogels met behulp van computervisie

Conclusie

Werken aan creatieve en experimentele projecten is een ongelooflijk voorrecht, het is wat mijn werk zo spannend maakt.

Met steeds groter wordende hoeveelheden data is digitale en autonome zingeving een grote uitdaging voor de toekomst: een waar ik graag aan werk!